JOURNAL OF MULTIVARIATE ANALYSIS 16, 54–70 (1985)

The Asymptotic Distribution of the Likelihood Ratio for Autoregressive Time Series with a Regression Trend*

ANDERS RYGH SWENSEN†

Norwegian Telecommunications Administration Research Establishment

Communicated by E. J. Hannan

It is shown that the likelihood ratio of an autoregressive time series of finite order with a regression trend is asymptotically normal. This result is used to derive the power of a test for positive correlation of the residuals under local autoregressive alternatives. The test is based on the Durbin–Watson statistics. © 1985 Academic Press, Inc.

1. INTRODUCTION

In this paper we shall deal with processes where an additional linear regression term is introduced. The asymptotic form of the likelihood ratio will be shown to be locally asymptotic normal (LAN) under rather weak distribution of assumptions on the distribution on the error terms. This fact makes it possible to find the asymptotic distribution of various statistics under contiguous alternatives, and hence compare them in terms of asymptotic efficiency and asymptotic power of tests.

We shall illustrate how this is done by finding the asymptotic power of a test for positive dependence between the error terms based on the Durbin–Watson statistics. Another possible application in the same vein,

Received January 18, 1982; revised April 22, 1983.

AMS 1980 subject classifications: primary 62M10; secondary 62E20, 62G10.

Key words and phrases: Regression, autoregressive errors, local asymptotic normality, test for dependence.

* This research was supported by the Norwegian Research Council for Science and the Humanities.

† Present address: Central Bureau of Statistics of Norway, P.O. Box 8131 Dep., N-0033 Oslo 1, Norway.

0047-259X/85 $3.00

which we would like to mention, is to the maximum likelihood estimator studied by Hannan, Dunsmuir, and Deistler [14]. They derived the estimator for a Gaussian model. However, asymptotically it is distribution-free under rather general conditions. Hence by applying Theorem 1, which states that the likelihood ratio is LAN, one should be able to compute the asymptotic distribution of this estimator over contiguous neighborhoods of the true parameter point for quite general error distributions.

For a detailed investigation of efficiencies in pure autoregressive processes, we refer to the paper by Akritas and Johnson [1].

Now, let us give some more details of the background for the paper. We shall first mention some results from asymptotic decision theory which are basic for the following. Suppose the observations $Y_1,..., Y_n$ have a distribution belonging to a parametric family $\{P_{\theta,n}: \theta \in \Theta\}$, where we only assume that Θ is an open set in a Euclidean space. Fix a particular point θ_0. The family $\{P_{\theta,n}: \theta \in \Theta\}$ is said to be locally asymptotic normal at θ_0, if there exists a sequence of random variables $\{Z_n\}$ and a positive definite matrix Γ so that for all t

$$\log(dP_{\theta_0+tn^{-1/2},n}/P_{\theta_0,n}) - t'Z_n + \tfrac{1}{2}t'\Gamma\, t \to 0 \tag{1.1}$$

in $P_{\theta_0,n}$ probability, and $Z_n \to N(0, \Gamma)$ in distribution under $P_{\theta_0,n}$. Here $dP_{\theta_0+tn^{-1/2},n}/dP_{\theta_0,n}$ denotes the Radon–Nikodym derivative of the part of the measure $P_{\theta_0+tn^{-1/2},n}$ which is absolutely continuous with respect to $P_{\theta_0,n}$.

The expansion (1.1) is exactly what is needed to be able to compute the efficiencies of estimates and tests under contiguous alternatives. But the implications are in fact much wider. From the theory due mainly to Le Cam and Hájek, it follows that it is possible to establish lower bounds for the limit of risks of rather general procedures and loss functions, and also to characterize the sequences that attain these lower bounds. For more details on this point one can consult the papers by Le Cam [18, 20] and Hájek [11, 12].

Furthermore, under the additional assumption that the LAN condition holds at every point $\theta \in \Theta$, Le Cam has showed how it is possible to construct estimators which are efficient in the sense that the lower bounds for the risks mentioned above, are obtained. The idea is to start out with a sequence which is $n^{1/2}$ consistent and modify it appropriately, see [17, 18, 20] for details.

Condition (1.1) and hence the corresponding results are valid under a variety of assumptions on the distributions of the observations $Y_1,..., Y_n$. Under some regularity conditions Le Cam has shown that (1.1) holds for independent, identically distributed observations, see [18, 19]. For observations from a stationary Markov chain, see Roussas [24]. Davies [5] and

Dzaparidze [9] have studied stationary Gaussian time series. Recently, also autoregressive time series with arbitrary error distribution have been treated by Akritas and Johnson [1]. Independent, nonidentically distributed random variables have been considered by Ibragimov and Khas'minski [16], and by Phillipou and Roussas [23]. For some more general types of dependence, see Roussas [25].

We want to point out that the rate of convergence $n^{-1/2}$ in (1.1) stems from the rate of separation of the sequences of measures $\{P_{\theta+tn^{-1/2},n}\}$ and $\{P_{\theta,n}\}$. In some cases other types of norming will give the correct rate of separation. This will be illustrated in the following.

In the remaining part of this section we shall state some auxillary results. The basic assumptions and main result of the paper will be given in Section 2. Section 3 contains the treatment of the test for positive dependence of the errors, while the last section contains the details of the proof of Theorem 2 of Section 2.

Let $P_{0,n}$ and $P_{1,n}$ be two sequences of probability measures on the measurable spaces $(\mathscr{X}_n, \mathscr{A}_n)$. Suppose that for each n there is a filtration $\mathscr{A}_{n,k} \subset \mathscr{A}_{n,k+1}$ of σ-algebras with $\mathscr{A}_{n,k_n} = \mathscr{A}_n$. Let $P_{0,n,k}$ and $P_{1,n,k}$ be the restrictions to $\mathscr{A}_{n,k}$ of $P_{0,n}$ and $P_{1,n}$, respectively. Let $\alpha_{n,k}$ be the Radon–Nikodym derivative on $\mathscr{A}_{n,k}$ of the part of $P_{1,n,k}$ which is dominated by $P_{0,n,k}$. Put $X_{n,k} = (\alpha_{n,k}/\alpha_{n,k-1})^{1/2} - 1$ where we take $\alpha_{n,0} = 1, n = 1,\dots$.

Using a truncation argument, and the central limit theorem for martingale differences due to McLeish [22], Le Cam [21] showed

THEOREM 1 (Le Cam). *Assume the following conditions are satisfied, all convergences beeing in probability under* $P_{0,n}$.

(i) $\max_k |X_{n,k}| \to 0$,

(ii) $\sum_k X_{n,k}^2 \to \tau^2/4$,

(iii) $\sum_k E(X_{n,k}^2 + 2X_{n,k} \mid \mathscr{A}_{n,k-1}) \to 0$,

(iv) $\sum_k E(X_{n,k}^2 I(X_{n,k} > 1) \mid \mathscr{A}_{n,k-1}) \to 0$.

Then the distribution of $\Lambda_n = \log dP_{1,n}/dP_{0,n}$ *is asymptotically equivalent to* $2\sum_k X_{n,k} - \tau^2/4$ *under* $P_{0,n}$, *and the distributions converge to the Gaussian distribution* $N(-\frac{1}{2}\tau^2, \tau^2)$.

Often, it will be possible to approximate the $X_{n,k}$'s with random variables $Z_{n,k}$, so that the following result is true.

LEMMA 1. *The following conditions are sufficient for* (i), (ii), *and* (iv) *of Theorem* 1 *to hold*:

$$E\sum_k (Z_{n,k} - X_{n,k})^2 \to 0, \tag{1.2}$$

$$\sup_n E\sum_k Z_{n,k}^2 < \infty, \tag{1.3}$$

$$\max_k |Z_{n,k}| \to 0, \tag{1.4}$$

$$\sum_k Z_{n,k}^2 \to \tau^2/4, \tag{1.5}$$

$$\sum_k E(Z_{n,k}^2 I(|Z_{n,k}| > \tfrac{1}{2}) | \mathscr{A}_{n,k-1}) \to 0, \tag{1.6}$$

where the convergence in (1.4)–(1.6) *are in probability under* $P_{0,n}$. *Furthermore, if condition* (iii) *of Theorem* 1 *is satisfied and*

$$E(Z_{n,k} | \mathscr{A}_{n,k-1}) = 0, \tag{1.7}$$

Λ_n *is asymptotically equivalent to* $2\sum_k Z_{n,k} - \tau^2/2$.

The first part of the lemma is immediate, while a proof of the second half can be found at the end of Section 4. Also, notice that condition (iii) of the theorem ensures that the singular part of $P_{1,n}$ with respect to $P_{0,n}$ on $\mathscr{A}_{n,k}$ does not behave too badly, When these measures are mutually absolutely continuous, this does not, of course, represent any problem.

It will thus be sufficient to verify conditions (1.2)–(1.7) to show that the autoregressive time series with a linear regression trend satisfy the LAN condition.

2. Assumptions and Main Results

Let $\{x_i\}_{i=1,\dots,}$ be a sequence of $q \times 1$ vectors, and suppose that the process $\{Y_n\}_{n=-p+1,-p+2,\dots,}$ can be written

$$Y_n = x_n'\beta + \eta_n, \qquad n = -p+1, \dots,$$

where

$$\eta_n = \varepsilon_n - \sum_{j=1}^{p} \theta_j \eta_{n-j}, \qquad n = 1, \dots,$$

with $x_i = 0$, $i \leqslant 0$. The vectors $\theta = (\theta_1, \dots, \theta_p)'$ and $\beta = (\beta_1, \dots, \beta_q)'$ are parameters. Let $X_n = \{x_{ij}\}$ be the matrix $(x_1 \vdots \cdots \vdots x_n)$ and define the $q \times q$ matrices,

$$A_n(m) = \{a_n(m)_{jk}\} = \sum_{i=1}^{n} x_i x_{i-m}', \qquad m = 0, \dots, p, n = m+1, \dots .$$

As usual, prime denotes transposition. We will make the following assumptions:

(A1) The solutions of the polynomial equation $1+\theta_1 z+\cdots+\theta_p z^p=0$ are all larger than 1 in modulus.

(A2) The random variables $\varepsilon_1, \varepsilon_2,\ldots,$ are independent and identically distributed with finite second moment σ^2 and expectation 0. Furthermore, their distribution is absolutely continuous with respect to Lebesgue measure, and the density f satisfies

(i) $f(z)>0$ for all z.

(ii) f is absolutely continuous with respect to Lebesgue measure.

(iii) If $\dot{f}$ is the derivative of f,

$$I(f)=\int(\dot{f}(z)/f(z))^2 f(z)\,dz<\infty.$$

Finally, we assume that

$$E(\phi(\varepsilon_1, f))^4<\infty, \qquad \text{where} \quad \phi(\cdot, f)=\dot{f}(\cdot)/f(\cdot).$$

(A3)(i) Let D_n be the diagonal matrix with elements $a_n(0)_{11},\ldots, a_n(0)_{qq}$, and put

$$R_n(m)=D_n^{-1/2}A_n(m)\,D_n^{-1/2}, \qquad m=0, 1,\ldots, p$$

Assume that $\lim R_n(m)=R(m)$, $m=0, 1,\ldots, p$, where $R(0)$ is positive definite.

(ii) $\max_k x_{ik}^2/\sum_{k=1}^n x_{ik}^2\to 0$ as $n\to\infty$, $i=1,\ldots, q$.

Comments on the Assumptions

(i) Assumptions (A1) and (A2) imply that the stochastic difference equation

$$\sum_{i=0}^{p}\theta_i\eta_{n-i}=\varepsilon_n, \qquad n=1, 2,\ldots,$$

where $\theta_0=1$, has a strictly stationary solution. It is not difficult, using arguments similar to those of Billingsley [3] and assumption (A2)(i), to show that the asymptotic expansion of the likelihood ratio will be the same for all solutions. Hence we can, without loss of generality, assume that we are dealing with the stationary one, so that $\{\eta_n\}_n$ is a pth order autoregressive process.

(ii) From a result in Hájek and Šidák [13, p. 212], see also Hájek

[12], it follows from (A2)(ii) and (iii) that $f^{1/2}$ is quadratic mean differentiable, i.e.,

$$\lim_{t\to 0} t^{-2} \int (f(z+t)^{1/2} - f(z)^{1/2} - \tfrac{1}{2} t\dot{f}(z)/f(z)^{1/2})^2 \, dz = 0$$

(iii) Assumptions (A3) concerning the behaviour of the regression coefficients are similar to those of Anderson [2]. We refer to this monograph for some discussion of the statistical implications.

(iv) The assumpion that $E(\phi(\varepsilon_1, f))^4 < \infty$ is superfluous in the case where the regression trend is absent.

Now let $\theta = (\theta_1,..., \theta_p)'$ and $\beta = (\beta_1,..., \beta_q)'$ be fixed, and let $t \in \mathbb{R}^p$, $b \in \mathbb{R}^q$. Denote the distribution of $(Y_{-p+1},..., Y_1,..., Y_n)$ under the parameter values (θ, β) by $P_{\theta,\beta,n}$. We want to find an asymptotic expression for the likelihood ratio

$$dP_{\theta + n^{-1/2}t, \beta + \sigma_n^{-1}b, n} / dP_{\theta,\beta,n}$$

where $\{\sigma_n\}$ is a suitably chosen sequence of $q \times q$ matrices.

Since $R(0)$ is assumed to be positive definite, there exists a $q \times q$ nonsingular matrix K so that $K'R(0)K = I$. Set $K_n = D_n^{-1/2}K$. Then it follows from the assumptions that

$$K_n' A_n(m) K_n = K' D_n^{-1/2} A_n(m) D_n^{-1/2} K = K' R_n(m) K \to K' R(m) K$$

for $m = 0, 1,..., p$. Now take $\sigma_n^{-1} = K_n$. Then we can write (with $t_0 = 0$)

$$\begin{aligned}
&\sum_{i=0}^{p} (\theta_i + n^{-1/2} t_i)(Y_{k-i} - x_{k-i}'(\beta + \sigma_n^{-1} b)) \\
&\quad = \sum_{i=0}^{p} \theta_i (Y_{k-i} - x_{k-i}'\beta) + n^{-1/2} \sum_{i=1}^{p} t_i (Y_{k-i} - x_{k-i}'\beta) \\
&\qquad - \sum_{i=0}^{p} \theta_i x_{k-i}' K_n b - n^{-1/2} \sum_{i=1}^{p} t_i x_{k-i}' K_n b \\
&\quad = \varepsilon_k + \gamma_{n,k}.
\end{aligned}$$

Referring to the random variables introduced in Theorem 1 and Lemma 1, we define

$$Z_{n,k} = 0, \qquad k = 1,..., p$$

$$Z_{n,k} = \frac{1}{2} \left(-\sum_{i=0}^{p} \theta_i x_{k-i}' K_n b + n^{-1/2} \sum_{i=1}^{p} t_i (Y_{k-i} - x_{k-i}'\beta) \right) \phi(\varepsilon_k, f),$$

$$k = p+1,..., n.$$

To apply the results from the Introduction we have to show that (1.2) is satisfied, i.e.,

$$\lim \sum_{k=1}^{n} E((f(\varepsilon_k+\gamma_{n,k})/f(\varepsilon_k))^{1/2}-1-Z_{n,k})^2=0.$$

This will follow from

$$\lim \sum_{k=1}^{n} E(Z_{n,k}-\tfrac{1}{2}\gamma_{n,k}\phi(\varepsilon_k,f))^2=0 \tag{2.1}$$

and

$$\lim \sum_{k=1}^{n} E((f(\varepsilon_k+\gamma_{n,k})/f(\varepsilon_k))^{1/2}-1-\tfrac{1}{2}\gamma_{n,k}\phi(\varepsilon_k,f))^2=0. \tag{2.2}$$

The details of the proof of (2.1) and (2.2) are given in the Appendix. Also the other conditions of Lemma 1 are shown to hold. In particular,

$$\sum_{k=1}^{n} Z_{n,k}^2 \to (I(f)\, b'K' \sum_{i,j} \theta_i R(|i-j|)\, \theta_j K)\, b/4 + I(f)\, t'\Delta t/4,$$

where Δ is the covariance matrix of $(\eta_1,\dots,\eta_p)$. We thus end up with

THEOREM 2. *Suppose the process* $\{Y_n\}_{n=1,2,\dots,}$ *satisfies Assumptions* (A1)–(A3). *Then*

$$\log dP_{\theta+n^{-1/2}t,\beta+K_n b,n}/dP_{\theta,\beta,n} - n^{-1/2} \sum_{k=1}^{n}\sum_{i=1}^{p} t_i\eta_{k-i}\phi(\varepsilon_k,f)$$
$$+\sum_{k=1}^{n}\sum_{i=0}^{p} \theta_i x'_{k-i}K_n b\phi(\varepsilon_k,f)+\tfrac{1}{2}(b' \vdots t')\Gamma_{\theta,\beta}(b' \vdots t')' \to 0$$

in $P_{\theta,\beta,n}$ *probability with* $K_n=D_n^{-1/2}K$, *where* $K'R(0)K=I$,

$$\Gamma_{\theta,\beta}=I(f)\left[\begin{array}{c|c} K'\sum_{i,j}\theta_i R(|i-j|)\,\theta_j K & 0 \\ \hline 0 & \Delta \end{array}\right]$$

and

$$\Delta=\begin{bmatrix} \Delta_0 & \cdots & \Delta_{p-1} \\ \vdots & & \vdots \\ \Delta_{p-1} & \cdots & \Delta_0 \end{bmatrix}$$

is the covariance matrix of $(\eta_1,\dots,\eta_p)$.

3. The Asymptotic Power of a Test for Positive Dependence

We shall now apply Theorem 1 to find the power under local first order autoregressive alternatives of a test for positive dependence based on the Durbins–Watson statistics. The procedure is straightforward. First we obtain a linear expansion of the statistics in question. Using this together with the expansion from Theorem 1, we get the joint asymptotic distribution of the likelihood ratio and the Durbin–Watson statistics. An application of Le Cam's third lemma, see Hájek and Šidák [13], yields the desired result. Assume the observations $Y_1,\dots, Y_n$ satisfy

$$Y_n = x_n'\beta + \eta_n, \qquad \eta_n = \theta\eta_{n-1} + \varepsilon_n,$$

where $|\theta| < 1$, and we let the assumptions otherwise be as in the previous section. In particular, we recall that there is no loss in generality to assume that $\eta_0, \eta_1,\dots,$ is a stationary sequence.

The least square estimator of β is

$$\hat{\beta}_n = (X_n X_n')^{-1} X_n (Y_1,\dots, Y_n)'$$

and the residuals are

$$\hat{\varepsilon}_i = Y_i - x_i'\hat{\beta}_n, \qquad i = 1,\dots, n.$$

The Durbin–Watson statistics, see [6–8], is defined as

$$S_n = \sum_{i=2}^{n} (\hat{\varepsilon}_i - \hat{\varepsilon}_{i-1})^2 \Big/ \sum_{i=1}^{n} \hat{\varepsilon}_i^2.$$

For testing the hypothesis that the errors are independent, i.e., $(\theta = 0, \beta = \beta)$, against the alternative that θ is positive, a reasonable approximate test is to reject if

$$\tfrac{1}{2} n^{1/2}(S_n - 2) > k_{1-\alpha}\sigma,$$

where $k_{1-\alpha}$ is the $1-\alpha$ fractile in the standard normal distribution. We shall now find the asymptotic power against alternatives of the form $(\theta = t/n^{1/2}, \beta = \beta)$.

As mentioned above, we first expand S_n under the hypotheses of independence, i.e., $\theta = 0$. To simplify the notation we suppress the index, n, writing X, $\hat{\beta}$, etc., instead of $X_n, \hat{\beta}_n$, etc. We remark that $\sum_{i=1}^{n} \hat{\varepsilon}_i^2$ may be written

$$\sum_{i=1}^{n} \varepsilon_i^2 - \varepsilon' X'(XX')^{-1} X\varepsilon,$$

where $\varepsilon' = (\varepsilon_1, ..., \varepsilon_n)$. Using the law of large numbers and assumption (A3), it follows that

$$n^{-1} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \to \sigma^2$$

in probability. Hence we only have to consider $n^{-1/2}(\sum_{i=2}^n (\hat{\varepsilon}_i - \hat{\varepsilon}_{i-1})^2 - 2\sum_{i=1}^n \hat{\varepsilon}_i^2)$ which by the stationarity of $\{Y_i - x_i'\beta\}_{i=1,2,...,}$ and the consistency of $\hat{\beta}$ is asymptotically equivalent to $-2n^{-1/2} \sum_{i=2}^n \hat{\varepsilon}_i \hat{\varepsilon}_{i-1}$.

However, taking $X_{-1} = (0 \vdots x_1 \vdots \cdots \vdots x_{n-1})$ and $X_1 = (x_2 \vdots \cdots \vdots x_n \vdots 0)$, one may write

$$\sum_{i=2}^{n} \hat{\varepsilon}_i \hat{\varepsilon}_{i-1} = \sum_{i=2}^{n} \varepsilon_i \varepsilon_{i-1} + \varepsilon' X'_{-1}(XX')^{-1} X\varepsilon$$
$$+ \varepsilon' X_1'(XX')^{-1} X\varepsilon + \sum_{i=2}^{n} x_i'(\hat{\beta} - \beta)\, x_{i-1}'(\hat{\beta} - \beta).$$

From assumption (A3) it follows that the last three terms on the right-hand side are tight. Hence

$$\tfrac{1}{2} n^{1/2}(S - 2) + (\sigma^2 n^{1/2})^{-1} \sum_{i=2}^{n} \varepsilon_i \varepsilon_{i-1} \to 0$$

in probability.

From Theorem 1 we obtain that the log likelihood ratio, Λ_n, in this case is asymptotically equivalent to

$$n^{-1/2} \sum_{i=1}^{n} t\varepsilon_{i-1} \phi(\varepsilon_i, f) - \tfrac{1}{2} t^2 \sigma^2 I(f).$$

Using the fact that the process $(\varepsilon_{i-1}\varepsilon_i, \varepsilon_{i-1}\phi(\varepsilon_i, f))$ is 1-dependent and stationary, it follows from Billingsley [4, p. 117], that $(\frac{1}{2}n^{1/2}(S-2), \Lambda_n)'$ asymptotically has a bivariate distribution with mean $(0, -\frac{1}{2}t^2\sigma^2 I(f))'$ and covariance matrix

$$\begin{bmatrix} 1 & -tE\varepsilon_1 \phi(\varepsilon_1, f) \\ -tE\varepsilon_1 \phi(\varepsilon_1, f) & t^2 I(f) \sigma^2 \end{bmatrix}$$

where the expectation is taken under the hypotheses $(\theta = 0, \beta = \beta)$. Therefore, from Le Cam's third lemma, cf. Hájek and Šidák [13], it follows that the asymptotic power of the test

$$n^{1/2}(S - 2) > 2k_{1-\alpha}\sigma$$

under alternatives ($\theta = t/n^{1/2}, \beta = \beta$) is

$$1 - \Phi(k_{1-\alpha} + tE_{0,\beta}\varepsilon_1\phi(\varepsilon_1, f)).$$

We remark that the asymptotic power is independent of β, hence it is the same as what we will get when testing for autoregressive dependence under the assumption that the observations are independent and identically distributed.

Also, we would like to point out that the expression above gives the asymptotic power of any test of the form

$$n^{1/2}(S-2) > 2k_{1-\alpha}\hat{\sigma}_n,$$

where $\hat{\sigma}_n$ is a consistent estimator for σ.

4. Appendix: Proof

We shall give here some details of the proof of Theorem 2. To do that we verify that the conditions of Lemma 1 and therefore also those of Theorem 1 are satisfied. The notation will be as in the previous sections with the exception that we let $\sum_k$ and $\sum_i$ mean $\sum_{k=1}^n$ and $\sum_{i=0}^p$, respectively.

Proof of (1.2). As pointed out in Section 2, this is done by verfying (2.1) and (2.2). As to (2.1), we remark that

$$\lim \sum_k E(Z_{n,k} - \tfrac{1}{2}\gamma_{n,k}\phi(\varepsilon_k, f))^2 = \tfrac{1}{4} \lim \sum_{k=1}^{p} E\gamma_{n,k}\phi(\varepsilon_k, f)^2 + \tfrac{1}{4} \lim n^{-1} I(f) \sum_k \left(\sum_i t_i x'_{k-i} K_n b \right)^2 = 0$$

since the first sum consists of only a finite number of terms which all tend to 0, and all elements of the vector $K'_n x_{k-i}$ are dominated in absolute value by

$$(x'_{k-i} K_n K'_n x_{k-i})^{1/2} = (x'_{k-i} D_n^{-1/2} KK' D_n^{-1/2} x_{k-i})^{1/2} \leqslant \left(\max_{i,j} \{K_{ij}^2\} \sum_{j=1}^{q} x_{j,k-i}^2 / a_n(0)_{jj} \right)^{1/2} \tag{4.1}$$

which tends to 0 by assumption (A3)(ii).

As to (2.2), it will be sufficient to show that $\lim(B_{1,n} + B_{2,n}) = 0$, where

$$B_{1,n} = \sum_k EI(|\eta_{k-i}| < K, i = 1,..., p)([f(\varepsilon_k + \gamma_{n,k})/f(\varepsilon_k)]^{1/2} - 1 - \tfrac{1}{2}\gamma_{n,k}\phi(\varepsilon_k, f))^2,$$

$$B_{2,n} = \sum_k EI(|\eta_{k-i}| \geqslant K, \text{ some } i = 1,\ldots, p)([f(\varepsilon_k + \gamma_{n,k})/f(\varepsilon_k)]^{1/2} - 1 - \tfrac{1}{2}\gamma_{n,k}\phi(\varepsilon_k, f))^2.$$

To show that $B_{1,n}$ and $B_{2,n}$ tends to 0 we will use the following result.

LEMMA 2. *Suppose that f satisfies* (A2)(i) *and* (ii). *Then for any sequence $v_n \to 0$*

$$\lim_n \sup_{|u| \leqslant b} \int ([f(y + un^{-1/2} + v_n)^{1/2} - f(y)^{1/2}]/(un^{-1/2} + v_n) - \tfrac{1}{2}\dot{f}(y)/f(y)^{1/2})^2\, dy = 0 \tag{4.2}$$

for all $b > 0$. Also

$$\int (f(y + v)^{1/2} - f(y)^{1/2} - \tfrac{1}{2}v\phi(y, f) f(y)^{1/2})^2\, dy \leqslant v^2 I(f). \tag{4.3}$$

Proof. Suppose (4.2) is not true. Then there is a sequence $\{u_n\}$ and u_0, $|u_0| \leqslant b$, so that $u_n \to u_0$, and

$$\int ([f(y + n^{-1/2}u_n + v_n)^{1/2} - f(y)^{1/2}]/(u_n n^{-1/2} + v_n) - \tfrac{1}{2}\dot{f}(y)/f(y)^{1/2})^2\, dy > \delta$$

for all n and some $\delta > 0$. But $u_n n^{-1/2} + v_n \to 0$ so this contradicts the assumption that $f^{1/2}$ is quadratic mean differentiable.

As to the inequality (4.3), we shall use the fact that (A2)(ii) and (A2)(iii), ensuring that f is absolutely continuous and has finite Fisher information, are sufficient for $f^{1/2}$ to be absolutely continuous, cfr Hájek and Šidák [13, p. 211]. Thus,

$$f(y)^{1/2} - f(y + v)^{1/2} = \tfrac{1}{2}\int_v^{y+v} \dot{f}(y)/f(y)^{1/2}\, dy.$$

From the Hölder inequality and the relation $(a + b)^2 \leqslant 2a^2 + 2b^2$, one now gets (4.3). ∎

Using (4.2) we see that $B_{1,n}$ is majorized by $c_n(K)\sum_k E\gamma_{n,k}^2$, where $c_n(K) \to 0$ as $n \to \infty$ for all K. But

$$\sum_k E\gamma_{n,k}^2 \leqslant 5\sum_k E(n^{-1/2}\sum_i t_i\eta_{k-i})^2 + 5\sum_k \left(\sum_i \theta_i x'_{k-i} K_n b\right)^2 + 5\sum_k (n^{-1/2}\sum_i t_i x'_{k-i} K_n b)^2. \tag{4.4}$$

The first term on the right-hand side in (4.4) stays bounded since by the ergodic theorem

$$n^{-1} \sum_k \eta_{k-i} \eta_{k-j} \to E\eta_1 \eta_{|i-j|} \qquad \text{a.s. and in } L_1.$$

A typical element in the second sum has the form

$$\sum_k \theta_i \theta_j (x'_{k-i} K_n b)(x'_{k-j} K_n b).$$

But by assumption (A3)(i)

$$\begin{aligned}
&\lim \sum_k (x'_{k-j} K_n b)(x'_{k-i} K_n b) \\
&\quad = \lim b' K'_n \sum_k x_{k-i} x'_{k-j} K_n b = \lim b' K'_n A_n(|i-j|)\, K_n b \\
&\quad = \lim b' K'\, D_n^{-1/2} A_n(|i-j|)\, D_n^{-1/2} K b = b' K' R(|i-j|)\, K b \qquad (4.5)
\end{aligned}$$

and thus the second term on the right-hand side in (4.4) stays bounded too. By the same argument the third term tends to 0. Thus, for any value of K

$$\lim B_{1,n} = 0.$$

Using the second part of the lemma

$$B_{2,n} \leqslant I(f) \sum_k \sum_i EI(|\eta_{k-i}| > K \text{ some } i = 1, \ldots, p)\, \gamma^2_{n,k}.$$

Bounding $\gamma^2_{n,k}$ as in (4.4), it follows by the stationarity of $\{\eta_n\}_{n=-p+1,\ldots}$, that $B_{2,n}$ can be made arbitrarily small uniformly in n by choosing K large enough.

Proof of (1.3). From the definition of $Z_{n,k}$ we have the following bound

$$\begin{aligned}
\sum_k EZ^2_{n,k} \leqslant 2I(f) \sum_k \Big(\sum_i x'_{k-i} K_n b \Big)^2 \\
+ 2n^{-1} E \sum_k \Big(\sum_i t_i \eta_{k-i} \phi(\varepsilon_k, f) \Big)^2
\end{aligned}$$

and it follows from (4.5) and the stationarity of $\{\eta_n\}_{n=-p+1,\ldots,}$ that $\sum_k EZ^2_{n,k}$ is uniformly bounded in n.

Proof of (1.4). We have to show that $\max_k |Z_{n,k}| \to 0$ in $P_{\theta,\beta,n}$ probability. This will follow from

$$\max_k \max_i |x'_{k-i} K_n b \phi(\varepsilon_k, f)| \to 0 \qquad (4.6)$$

and

$$\max_k \max_i |n^{-1/2}\eta_{k-i}\phi(\varepsilon_k, f)| \to 0 \tag{4.7}$$

both in $P_{\theta,\beta,n}$ probability.

But $\phi(\varepsilon_k, f)$, $k = 1,\ldots,$ are i.i.d. with finite second moment and zero expectation. Hence (4.6) will follow from (4.1) and (4.5) since $P_{\theta,\beta,n}(\max_k \max_i |x'_{k-i}K_n b\phi(\varepsilon_k, f)| > \delta)$ is majorized by

$$\delta^{-2} \sum_k \sum_i (x'_{k-i}K_n b)^2 E\phi(\varepsilon_k, f)^2 I(|x'_{k-i}K_n b\phi(\varepsilon_k, f)| > \delta).$$

The relation (4.7) is proved in a similar way using the stationarity of $\{\eta_i\}_{i=-p+1,\ldots}$.

Proof of (1.5). We will show that $\sum_k Z^2_{n,k}$ converges in probability under $P_{\theta,\beta,n}$. But up to a term which tends to 0 in $P_{\theta,\beta,n}$ probability when $n \to \infty$

$$\begin{aligned}\sum_k Z^2_{n,k} &= \sum_k \left(\sum_i \theta_i x'_{k-i} K_n b\right)^2 \phi(\varepsilon_k, f)^2/4 \\ &\quad + n^{-1} \sum_k \left(\sum_i t_i \eta_{k-i}\right)^2 \phi(\varepsilon_k, f)^2/4 \\ &\quad - \sum_k n^{-1/2} \left(\sum_i \theta_i x'_{k-i} K_n b\right)\left(\sum_i t_i \eta_{k-i}\right) \phi(\varepsilon_k, f)^2/2.\end{aligned}$$

Now from (4.1) and (4.5) it follows that

$$\max_k \left(\sum_i \theta_i x'_{k-i} K_n b\right)^2 \left(\sum_k \left(\sum_i \theta_i x'_{k-i} K_n b\right)^2\right)^{-1} \to 0.$$

Thus $\sum_k \sum_i \theta_i x'_{k-i} K_n b\phi(\varepsilon_k, f)$ is asymptotically normal, and by relative stability, cf. Gnedenko and Kolmogorov [10],

$$\sum_k \left(\sum_i \theta_i x'_{k-i} K_n b\right)^2 \phi(\varepsilon_k, f)^2 \to I(f)\, b' \sum_{i,j} \theta_i K' R(|i-j|) K\theta_j b \tag{4.8}$$

in $P_{\theta,\beta,n}$ probability.

By the ergodic theorem

$$n^{-1} \sum_k \left(\sum_i t_i \eta_{k-i}\right)^2 \phi(\varepsilon_k, f)^2 \to I(f)\, t' \Delta t,$$

where Δ is the covariance matrix of $(\eta_1,\ldots, \eta_p)'$. Finally, that the cross term tends to 0 in $P_{\theta,\beta,n}$ probability will follow from the Markov inequality if we show that

$$\lim n^{-1}E\left(\sum_k\left(\sum_i \theta_i x'_{k-i}K_n b\right)\left(\sum_i t_i\eta_{k-i}\right)\phi(\varepsilon_k,f)^2\right)^2=0.$$

But this is a consequence of assumption (A2)(iii), (4.1), the fact that $\{\eta_n\}_{n=-p+1,\ldots,}$ is stationary and the fact that $E|\eta_1\eta_j|<(\text{const.})\,d^j$, where $0<d<1$. Hence (1.5) is satisfied with τ^2 equal to

$$I(f)b'\left(K'\sum_{i,j}\theta_i R(|i-j|)\,\theta_j K\right)b+I(f)\,t'\,\Delta t$$

Proof of (1.6). Remark that

$$E[EZ^2_{n,k}I(|Z_{n,k}|>\tfrac{1}{2})\mid \mathscr{A}_{n,k-1}]$$

is bounded by

$$EZ^2_{n,k}I(\max_k|Z_{n,k}|>\tfrac{1}{2}).$$

Hence it will be sufficient to show that $\sum_k Z^2_{n,k}$ is uniformly integrable. However,

$$Z^2_{n,k}\leqslant 2\left(\sum_i \theta_i x'_{k-i}K_n b\phi(\varepsilon_k,f)\right)^2+2\left(\sum_i t_i\eta_{k-i}\phi(\varepsilon_k,f)\right)^2. \tag{4.9}$$

From (4.8) and the convergence of the sum of the means of the first term on the right-hand side of (4.9), it follows that this sum is uniformly integrable. The same is true for the second sum by the L_1 convergence in the ergodic theorem.

Proof of (1.7). From the definition of $Z_{n,k}$ it follows that (1.7) is satisfied if $E\phi(\varepsilon_k,f)=0$, $k=1,2,\ldots$, *i.e.*, if $\int \dot f(z)\,dz=0$. But this is a consequence of Lemma I.2.4.a in Hájek and Šidák [13].

Proof of Lemma 1. The first half of the lemma is straightforward, so we only prove the second half. Take $V_{n,k}=X_{n,k}-Z_{n,k}-E(X_{n,k}\mid\mathscr{A}_{n,k-1})$. Then $E(V_{n,k}\mid\mathscr{A}_{n,k-1})=0$. By a well-known martingale inequality

$$P_{0,n}\left(\left|\sum_k V_{n,k}\right|^2>\delta\right)\leqslant\delta^2 E\sum_k V^2_{n,k}.$$

But

$$E\sum_k V^2_{n,k}\leqslant 2E\sum_k(X_{n,k}-Z_{n,k})^2+2E\sum_k(E(X_{n,k}\mid\mathscr{A}_{n,k-1}))^2.$$

The first term on the right-hand side tends on 0 by assumption (1.2), while the last one satisfies

$$E\sum_k (E(X_{n,k}\,|\,\mathscr{A}_{n,k-1}))^2 \leqslant 2E\sum_k (X_{n,k}-Z_{n,k})^2 + 2E\sum_k (E(Z_{n,k}\,|\,\mathscr{A}_{n,k-1}))^2$$

which tends to 0 by (1.2) and (1.7). We shall now show that

$$\sum_k E(X_{n,k}\,|\,\mathscr{A}_{n,k-1}) \to -\tau^2/8 \tag{4.10}$$

in $P_{0,n}$ probability. This will prove the lemma, since then

$$\sum_k X_{n,k} - Z_{n,k} + \tau^2/8 \to 0$$

in $P_{0,n}$ probability. The statement (iii) of Theorem 1 implies that (4.10) is equivalent to

$$\sum_k E(X_{n,k}^2\,|\,\mathscr{A}_{n,k-1}) \to \tau^2/4$$

in $P_{0,n}$ probability. But by (1.2) and (1.6) this is in turn equivalent to

$$\sum_k E(Z_{n,k}^2 I(|Z_{n,k}| \leqslant \tfrac{1}{2})\,|\,\mathscr{A}_{n,k-1}) \to \tau^2/4.$$

Since $P_{0,n}(\sum_k Z_{n,k}^2 I(|Z_{n,k}| > \tfrac{1}{2}) > \delta) \leqslant P_{0,n}(\max_k |Z_{n,k}| > \delta)$ when $0 < \delta < \tfrac{1}{2}$, it follows from (1.4) and (1.5) that

$$\sum_k Z_{n,k}^2 I(|Z_{n,k}| \leqslant \tfrac{1}{2}) \to \tau^2/4$$

in $P_{0,n}$ probability. Hence (4.10) will follow if we show that

$$\sum_k Z_{n,k}^2 I(|Z_{n,k}| \leqslant \tfrac{1}{2}) - E(Z_{n,k}^2 I(|Z_{n,k}| \leqslant \tfrac{1}{2})\,|\,\mathscr{A}_{n,k-1}) \to 0 \tag{4.11}$$

in $P_{0,n}$ probability.

But here we can use some results from the literature on dependent central limit theorems. By (3.15) of McLeish [22] it follows that the conditional Lindeberg condition

$$\sum_k E(Z_{n,k}^2 I(\tfrac{1}{2} \geqslant |Z_{n,k}| > \delta\,|\,\mathscr{A}_{n,k-1}) \to 0 \tag{4.12}$$

in $P_{0,n}$ probability for all $\delta > 0$, and

$$\lim_{K\to\infty} \limsup_n P_{0,n}\left(\sum_k E(Z_{n,k}^2 I(|Z_{n,k}| \leqslant \tfrac{1}{2})\,|\,\mathscr{A}_{n,k-1}) > K\right) = 0 \tag{4.13}$$

are sufficient for (4.11) to hold. From (1.3) it is immediate that (4.13) is satisfied. Finally the conditional Lindeberg condition (4.12) follows by an argument along the lines of Lemma 3.5 in Helland [15].

ACKNOWLEDGMENTS

Most of the material in this paper is from the authors's Ph. D. dissertation, University of California, Berkeley, 1980, written under the supervision of Professor Lucien Le Cam, to whom I want to express my gratitude for valuable help and advice. Also, I would like to thank the referees for some valuable comments that improved the quality of the paper.

REFERENCES

[1] AKRITAS, M. G. AND JOHNSON, R. A. (1980). Efficiencies of tests and estimators for p-order autoregressive processes when error distribution is nonnormal. Unpublished manuscript, University of Wisconsin, Madison.

[2] ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.

[3] BILLINGSLEY, P. (1961). *Statistical Inference for Markov Processes.* University of Chicago Press, Chicago.

[4] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[5] DAVIES, R. B. (1973). Asymptotic inference in stationary Gaussian time series. *Advan. Appl. Probab.* **5** 469–497.

[6] DURBIN, J. AND WATSON, G. S. (1950). Testing for serial correlation in least squares regression I. *Biometrika* **37** 409–428.

[7] DURBIN, J. AND WATSON, G. S. (1951). Testing for serial correlation in least squares regression II. *Biometrika* **38** 159–178.

[8] DURBIN, J. AND WATSON, G. S. (1971). Testing for serial correlation in least squares regression III. *Biometrika* **58** 1–19.

[9] DZHAPARIDZE, K. O. (1977). Estimation of parameters of a spectral density with fixed zeroes. *Theory Probab. Appl.* **22** 708–729.

[10] GNEDENKO, B. V. AND KOLMOGOROV, A. N. (1954). *Limiting Distributions for Sums of Independent Random Variables.* Addison–Wesley, Reading, Mass.

[11] HÁJEK, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. Verw. Gebiete* **14** 323–330.

[12] HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I, pp. 175–194, University of California Press, Berkeley and Los Angeles.

[13] HÁJEK, J. AND ŠIDÁK, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.

[14] HANNAN, E. J., DUNSMUIR, W. T. M. AND DEISTLER, M. (1980). Estimation of vector Armax models. *J. Multivariate Anal.* **10** 275–295.

[15] HELLAND, I. S. (1980). On weak convergence to Brownian motion. *Z. Wahrsch. Verw. Gebiete* **52** 251–265.

[16] IBRAGIMOV, I. A. AND KHAS'MINSKII, R. Z. (1975). Local asymptotic normality for non-identically distributed observations. *Theory Probab. Appl.* **20** 246–260.

[17] LE CAM, L. (1956). On the asymptotic theory of estimation and testing hypothesis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I, pp. 129–156, Univ. of California Press, Berkeley and Los Angeles.

[18] LE CAM, L. (1960). Locally asymptotically normal families of distributions. In *University of California Publ. in Statistics*, Vol. 3, pp. 27–98.
[19] LE CAM, L. (1966). Likelihood functions for large numbers of independent observations. In *Research Papers in Statistics* (F. N. David, Ed.), pp. 167–187, Wiley, New York.
[20] LE CAM, L. (1974). *Notes on Asymptotic Methods in Statistical Decision Theory.* Centre de Recherches Mathématiques, Université de Montréal.
[21] LE CAM, L. (1980). *Asymptotic Methods in Statistical Decision Theory.* In preparation.
[22] MCLEISH, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **2**, 620–628.
[23] PHILIPPOU, A. N. AND ROUSSAS, G. G. (1973). Asymptotic distribution of the likelihood function in the independent not identically distributed case. *Ann. Statist.* **1** 454–471.
[24] ROUSSAS, G. G. (1972). *Contiguity of Probability Measures: Some Applications in Statistics.* Cambridge Univ. Press, London.
[25] ROUSSAS, G. G. (1979). Asymptotic distribution of the log likelihood function for stochastic processes. *Z. Wahrsch. Verw. Gebiete* **47**, 31–46.